AMENDMENTS TO THE SPECIFICATION:

Please replace the paragraph beginning at line 16 on page 8 with the following:

To be specific, in FIG. 3A, a video sequence and query images (1 through n) are input (step 300). The video sequence is divided into one or more shots (1 through p), a key frame is selected in each of the shots (step 302). Next, p key frames are buffered (step 304). FIG. 4 shows an example of a video sequence divided into p shots and their key frames. In FIG. 4, the first frame is selected from each of the shots as key frames ~~as key frames KF!1, KF!2, KF!3, . . . , and KF!p~~ <u>KF-1, KF-2, KF-3, . . . , and KF-p</u>. FIG. 5 shows an example of dividing a video sequence extracted from a soap opera into 8 shots and selecting their key frames. In FIG. 5, the video sequence consisting of 619 frames in total are divided into 9 shots, and the key frame of each of the shots is designated by frame number.

Please replace the paragraph beginning at line 13 on page 9 with the following:

To be specific, ~~nHp~~ <u>n x p</u> shot mask images are created with respect to p key frames and n query objects (the pixels of the shot mask image without the object region are all turned off ("0")).